

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 18, 2003, 03:18:06 ; Search time 9.44262 Seconds
(without alignments)
1581.285 Million cell updates/sec

Title: US-09-807-933B-5

Perfect score: 1956

Sequence: 1 MKFLTIASSAILALAVGTEM.....TYKQVTCPKAITAKSGCSRK 360

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	722.5	36.9	376	1 GUNK_FUSOX	P45699 fusarium ox
2	675.5	34.5	213	1 GUN5_HUMIN	P43316 humicola in
3	524.5	26.8	511	1 GUNB_PSEFL	P18126 pseudomonas
4	490	25.1	393	1 GUN1_USTWA	P54424 ustilago ma
5	226	11.6	471	1 GUX2_TRIRE	P07987 trichoderma
6	198.5	10.1	475	1 SIM1_YEAST	P40472 saccharomyc
7	196	10.0	662	1 MUC1_XENLA	Q05049 xenopus lae
8	172.5	8.8	462	1 GUNB_FUSOX	P46236 fusarium ox
9	169.5	8.7	389	1 SER1_BOMMO	P07856 bombyx mori
10	169	8.6	567	1 CH13_CANAL	P40954 candida alb
11	164	8.4	797	1 VGLX_HSVB	P28968 equine herp
12	162.5	8.3	418	1 GUN2_TRIRE	P07982 trichoderma
13	162.5	8.3	1322	1 YAG3_YEAST	P39712 saccharomyc
14	162	8.3	3178	1 YS89_CAEEL	Q09624 caenorhabdi
15	160.5	8.2	786	1 STUB_DROME	Q05319 drosophila
16	160	8.2	503	1 YN23_YEAST	P38332 saccharomyc
17	158	8.1	1161	1 DAN4_YEAST	P47179 saccharomyc
18	156.5	8.0	263	1 SGS3_DROXA	P13728 drosophila
19	156	8.0	438	1 GUX3_AGABI	P49075 agaricus bi
20	155.5	7.9	1251	1 YQ03_CAEEL	Q09550 caenorhabdi
21	155	7.9	217	1 SGS3_DROSI	P13729 drosophila
22	153.5	7.8	725	1 AGA1_YEAST	P32323 saccharomyc
23	153	7.8	299	1 YJH9_YEAST	P47032 saccharomyc
24	149.5	7.6	210	1 PSBP_PORPU	P50272 porphyra pu
25	149	7.6	605	1 YHC8_YEAST	P38739 saccharomyc
26	147.5	7.5	420	1 SUN4_YEAST	P53616 saccharomyc
27	143.5	7.3	624	1 YH19_YEAST	P38900 saccharomyc
28	143.5	7.3	881	1 YJH8_YEAST	P47033 saccharomyc
29	143	7.3	1367	1 AMYH_YEAST	P08640 saccharomyc
30	141.5	7.2	748	1 GUNC_PSEFL	P27033 pseudomonas
31	140.5	7.2	1140	1 YW96_YEAST	Q04893 saccharomyc
32	139.5	7.1	1374	1 YWN3_YEAST	Q03099 saccharomyc
33	139	7.1	1075	1 FLO5_YEAST	P38894 saccharomyc

ALIGNMENTS

RESULT 1

ID	GUNK_FUSOX	STANDARD;	PRT;	376 AA.
AC	P45699;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-NOV-1995 (Rel. 32, Last annotation update)			
DE	Putative endoglucanase type K precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase) (Cellulase).			
OS	Fusarium oxysporum.			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OC	Hypocreales; mitosporic Hypocreales; Fusarium.			
OX	NCBI_Taxid=5507;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=95047531; PubMed=7959045;			
RA	Sheppard P.O., Grant F.J., Oort P.J., Sprecher C.A., Foster D.C.,			
RA	Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;			
RT	"The use of conserved cellulase family-specific sequences to clone			
RT	cellulase homologue cDNAs from Fusarium oxysporum.";			
RL	Gene 150:163-167(1994).			
CC	-!- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic			
CC	linkages in cellulose.			
CC	-!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).			
CC	-!- SIMILARITY: BELONGS TO CELLULOSE FAMILY K (FAMILY 45 OF GLYCOSYL			
CC	HYDROLASES).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; L29381; AAA65589.1; --			
DR	HSSP; P43316; 2ENG.			
DR	InterPro; IPR000254; CBD_fungal.			
DR	InterPro; IPR000334; GH_45.			
DR	Pfam; PF00734; CBM 1; 1-			
DR	Pfam; PF02015; Glyco_hydro_45; 1.			
DR	SMART; SM00236; fCBD; 1.			
DR	PROSITE; PS00562; CBD_FUNGAL; 1.			
DR	PROSITE; PS01140; GLYCOSYL_HYDROL_F45; 1.			
KW	Cellulose degradation; Hydrolase; Glycosidase; Signal.			
FT	SIGNAL 1 18			POTENTIAL.
FT	CHAIN 19 376			PUTATIVE ENDOGLUCANASE TYPE K.
FT	DOMAIN 19 308			CATALYTIC.
FT	DOMAIN 309 338			LINKER.
FT	DOMAIN 339 376			CELLULOSE-BINDING.
FT	ACT SITE 29 29			NUCLEOPHILE (BY SIMILARITY).
FT	ACT SITE 140 140			PROTON DONOR (BY SIMILARITY).
SQ	SEQUENCE 376 AA; 39235 MW; B430A5F962B9F882 CRC64;			

Query Match 36.9%; Score 722.5; DB 1; Length 376;
Best Local Similarity 59.8%; Pred. No. 6e-40;

Matches 128; Conservative 31; Mismatches 48; Indels 7; Gaps 4;
 QY 147 AVSGASGNGETRYWDCCKPSCSWPGKADVTSPVSGNCKDKGKTLADNNTQNCV-GGSS 205
 DB 14 AVS-AASGSHSTRYWDCCKPSCSWGSKAAVAPALATCDKNDNPISNTNVAVCEGGGSA 72
 QY 206 YTCNDNPQWVSDLLAYGFAAASISGGSEATWCCACFELTFTSTAVKGMVQVNTNTGS 265
 DB 73 YACTNYSFWANDELAYGFAATKISGGSEASWCCACVATFTTGPVKGKMIQVSTNTGG 132
 QY 266 DLGNTGAGHDLQMPGGVGIYNGCATOWGAPTDCWGARYGVSSASDCSNLPSALQAGC 325
 DB 133 DLGDN---HFDLMPGGVGIYNGCATOWGAPTDCWGARYGVSSASDCSNLPSALQAGC 187
 QY 326 KWRFGWFKNADNPTMYKQVTCPKAITAKSGCSR 359
 DB 188 HWRFDWFENADNPTMYKQVTCPKAITAKSGCSR 221

RESULT 2
 GUNB_HUMIN STANDARD; PRT; 213 AA.
 AC P43316;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Endoglucanase V (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
 DE (Cellulase V) (EG V).
 OS Humicola insolens.
 OC Eukaryota; Fungi; Ascomycota; mitosporic Ascomycota; Humicola.
 OX NCBI_TaxID=34413;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rasmussen G., Mikkelsen J.-M., Schulein M., Patkar S.A., Hagen F.,
 Hjort C.M., Hastrup S.;
 RA "A cellulase preparation comprising an endoglucanase enzyme."
 RT Patent number WO9117243, 14-NOV-1991.
 RL [2]
 RN X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
 RX MEDLINE=93390621; PubMed=8377830;
 RA Davies G.J., Dodson G.G., Hubbard R.E., Tolley S.P., Dauter Z.,
 Wilson K.S., Hjort C., Mikkelsen J.M., Rasmussen G., Schulein M.;
 RA "Structure and function of endoglucanase V."
 RT Nature 365:362-364 (1993).
 RL [3]
 RN X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
 RX MEDLINE=96101453; PubMed=8519779;
 RA Davies G.J., Tolley S.P., Henriessat B., Hjort C., Schulein M.;
 RA "Structures of oligosaccharide-bound forms of the endoglucanase V
 from Humicola insolens at 1.9-A resolution."
 RT Biochemistry 34:16210-16220 (1995).
 RL [4]
 RN X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
 RA Davies G.J., Dodson G.G., Moore M.H., Tolley S.P., Dauter Z.,
 Wilson K.S., Rasmussen G., Schulein M.;
 RA "Structure determination and refinement of the Humicola insolens
 endoglucanase V at 1.5-A resolution."
 RT Acta Crystallogr. D 52:7-17 (1996).
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 linkages in cellulose.
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL
 HYDROLASES).
 CC PDB; 2ENG; 08-DEC-96.
 CC PDB; 3ENG; 16-JUN-97.
 CC PDB; 4ENG; 16-JUN-97.
 CC InterPro: IPR000334; GH_45.
 CC Pfam: PF02015; Glyco_hydro_45; 1.
 CC PROSITE: PS01140; GLYCOSYL_HYDROL_F45; 1.
 CC Cellulose degradation; Hydrolase; Glycosidase; 3D-structure.
 KW ACT_SITE 10 10 NUCLEOPHILE.
 FT ACT_SITE 121 121 PROTON DONOR.
 FT SEQUENCE 213 AA; 22864 MW; 24334301BA3BC804 CRC64;
 SQ

Query Match 34.5%; Score 675.5; DB 1; Length 213;
 Best Local Similarity 57.3%; Pred. No. 3.7e-37; Indels 7; Gaps 3;
 Matches 118; Conservative
 QY 155 NGETRYWDCCKPSCSWPGKADVTSPVSGNCKDKGKTLADNNTQNCV-GGSSYTCNDNP 213
 DB 2 DGRSTRYWDCCKPSCSWGAKAPVNPVFCNANFORITDFDAKSGCEPGGVAYSCADQTP 61
 QY 214 WVSDLLAYGFAAASISGGSEATWCCACFELTFTSTAVKGMVQVNTNTGSDLGNTGA 273
 DB 62 WVNDDFALGFAATSIAGSEAGWCCACVATFTTSGVAGKMWVQSTSTGGDLGNS--- 118
 QY 274 HFDLMPGGVGIYNGCATOWGAPTDCWGARYGVSSASDCSNLPSALQAGCKWRFGWFK 333
 DB 119 HFDLNPGGVGIYNGCATOWGAPTDCWGARYGVSSASDCSNLPSALQAGCKWRFGWFK 175
 QY 334 NADNPTMYKQVTCPKAITAKSGCSR 359
 DB 176 NADNPTMYKQVTCPKAITAKSGCSR 201

RESULT 3
 GUNB_PSEFL STANDARD; PRT; 511 AA.
 AC P18126;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Endoglucanase B precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
 DE (Cellulase) (EGB).
 GN CELB.
 OS Pseudomonas fluorescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=294;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 30-48.
 RX STRAIN=Sp. Cellulosa;
 RX MEDLINE=50355836; PubMed=2117693;
 RA Gilbert H.J., Hall J., Hazlewood G.P., Ferreira L.M.A.;
 RA "The N-terminal region of an endoglucanase from Pseudomonas
 fluorescens subspecies cellulosa constitutes a cellulose-binding
 domain that is distinct from the catalytic centre."
 RT Mol. Microbiol. 4:759-767 (1990).
 CC -1- FUNCTION: THIS ENZYME CATALYZES THE ENDOHYDROLYSIS OF 1,4-BETA-
 GLUCOSIDIC LINKAGES IN CELLULOSE, LICHENIN AND CEREAL BETA-D-
 GLUCANS. EGB IS MOST ACTIVE AGAINST BARLEY BETA-GLUCAN, BUT SHOWED
 SIGNIFICANT ACTIVITY AGAINST AMORPHOUS AND CRISTALLINE CELLULOSE.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
 linkages in cellulose.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
 (CBD).
 CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL
 HYDROLASES).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC EMBL; X52615; CAA36844.1;
 CC PIR; S10527; S10527.
 CC HSSP; P43316; 2ENG.
 CC InterPro: IPR001919; Bac celose-bind.
 CC InterPro: IPR002893; CBD_5.
 CC InterPro: IPR000334; GH_45.
 CC Pfam; PF00553; CBM_2; 1.
 CC Pfam; PF02013; CBM_10; 1.
 CC Pfam; PF02015; Glyco_hydro_45; 1.

DR PROSITE; PS00561; CBD BACTERIAL; 1.
DR PROSITE; PS01140; GLYCOSYL HYDROL F45; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Periplasmic.
FT SIGNAL 1 29
FT CHAIN 30 511
FT DOMAIN 30 131
FT DOMAIN 132 173
FT DOMAIN 223 259
FT DISULFID 32 127
FT ACT_SITE 276 276
FT ACT_SITE 393 393
SQ SEQUENCE 511 AA; 52078 MW; 3C3119D998291D8E CRC64;

Query Match 26.8%; Score 524.5; DB 1; Length 511;
Best Local Similarity 30.8%; Pred. No. 4.5e-27;
Matches 123; Conservative 59; Mismatches 119; Indels 99; Gaps 12;

Qy 33 CGGKNDGPTCCSGSTCVDPDNPFFYSQCPVNPENLSTNKSHTTTESAKITTTKG 92
Db 127 CGGQSGAPSSVASSSS-----SSSVSSSTPRSSSSSVSVPGTSSSS 171
Qy 93 SKKTTTTEASK-----KTT-----TTEASKTT 115
Db 172 SSVLTGAQCNWYGLTLPCLNNTNGWYEDGRCVARTTCSAQPAPYGVISTSSSTPLS 231
Qy 116 TTEASKTTTTTTHKASTSSSSSASTNYSAGVGGASGNGETTRYWDCKPSCSWPCKA 175
Db 232 SSSSRSSVASSSSLSASTSSASVS-SVPPIDGGC--NGYATRYWDCKPCHGMSANV 288
Qy 176 -DVTSVPGSCNKGKTLADNNTQNGCVGSSYTCNDNQPVWVSDDLAYGFAAASISGSGE 234
Db 289 PSLVPLQSCANTRLSVSGSCDGGGYMCWDKIPPAVSPTLAYGYAATS-SGDV- 346
Qy 235 ATWCCACPELTFT-----STAVKGMVQVNTGSDLSNTGAHFDLQMPGG 283
Db 347 ---CGRCYQLQFTGSSYNAPDPSAALAGKTMIVQATNIGYDV---SGGQFDILVPGG 400
Qy 284 VGIYNGCATONGAPTGDGNGARYGG-----VSSASDC-----SNLPS 319
Db 401 VGAFNACSAQNGVNAELGAQYGFGLAACKOOLGYNASLSQKCVLNRCDSVFGSRLT 460
Qy 320 ALQAGCKWRFGWKNADNPTMYQVTCPCPKAITAKSGCSR 359
Db 461 QLQCGCTWFASWFEAADNPSLKYKEVPCPAELTTRSGMNR 500

RESULT 4
GUNI_GUN1
ID -GUNI_US1TWA STANDARD; PRT; 393 AA.
AC P54424;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Endoglucanase 1 precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase 1)
DE (Cellulase 1) (EG 1).
GN EG1.
OS Ustilago maydis (Smut fungus).
OC Eukaryota; Fungi; Basidiomycota; Ustilaginomycetes;
OC Ustilaginomycetidae; Ustilaginales; Ustilaginaceae; Ustilago.
OX NCBI_TaxID=5270;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FB11.
RX MEDLINE=96145728; PubMed=8590631;
RA Schaeuwer F., Warner G., Kahmann R.;
RT "Flament-specific expression of a cellulase gene in the dimorphic
RT fungus Ustilago maydis";
RL Biol. Chem. Hoppe-Seyler 376:617-625(1995).
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: HYPHAL TIP.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN FILAMENTOUS DIKARYON.

CC -1- PTM: MAY ALSO BE O-GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY K (FAMILY 45 OF GLYCOSYL
CC HYDROLASES).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S81598; AAB36147.1; -.
DR HSPP; P43316; 2ENG.
DR InterPro; IPR000334; GE 45.
DR Pfam; PF02015; Glyco_hydro_45; 1.
DR PROSITE; PS01140; GLYCOSYL-HYDROL F45; 1.
KW Cellulose degradation; Hydrolase; Glycosidase; Signal; Glycoprotein.
FT SIGNAL 1 26
FT CHAIN 27 393
FT ACT_SITE 34 34
FT ACT_SITE 152 152
FT DOMAIN 270 385
FT DOMAIN 343 343
FT CARBOHYD N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 393 AA; 39594 MW; 65C753C610CD6AD3 CRC64;

Query Match 25.1%; Score 490; DB 1; Length 393;
Best Local Similarity 46.1%; Pred. No. 5.8e-25;
Matches 101; Conservative 29; Mismatches 67; Indels 22; Gaps 9;

Qy 156 GETTRYWDCKPSCSWPFGKADVTSPVSGCNKGKTLADN-----NTQNGCVGSSYTCNDN 211
Db 27 GMATRYWDCLASASWEGKAPVAVPADKADGVTLLDSKDPGSGQSGCNGKFMCSM 86
Qy 212 QPWWVSD--LAYGFAAASISGSGSEATWCCACPELTFT-----TSTAVKGMVQVNTGSG 265
Db 87 QPFDEDTPTLAFGFGA--FTTQGESDTCACFYAEFEHDAQGMKRNKLIQVNTVGG 144
Qy 266 DLGSNTGAHFDLQMPGGVGIY-NGCATONGAPTGDGNGARYGVSSASDCSNLPSALQAG 324
Db 145 DVQSQ---NFDQFQGGGLGAFPKGCPAQGVASLWGDQYGVKSAITECKLPKPLQEG 201
Qy 325 CKWRFG-WFKNADNPTM--TYQVTCPCPKAITAKSGCSRK 360
Db 202 CKWRFGSEW---GDNPLKSGPKRVKCPKSLIDRSGCQRK 237

RESULT 5
GUX2_TRIE
ID -GUX2_TRIE STANDARD; PRT; 471 AA.
AC P07987;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Exoglucanase II precursor (EC 3.2.1.91) (Exocellobiohydrolase II)
DE (CBHII) (1,4-beta-cellobiohydrolase).
GN CBH2.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VT-D-80133;
RX MEDLINE=87248061; PubMed=3596237;
RA Teeri T.T., Lehtovaara P., Kauppinen S., Salovuori I., Knowles J.;
RT "Homologous domains in Trichoderma reesei cellulolytic enzymes: gene
RT sequence and expression of cellobiohydrolase II";
RL Gene 51:43-52(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=QM9414 / Rut C-30;
RA Chen C.M., Gritzali M., Stafford D.W.;

RT "Nucleotide sequence and deduced primary structure of
 RT cellobiohydrolase II from *Trichoderma reesei*.";
 RL Biotechnology 5:274-278 (1987).
 RN [3]
 RP SEQUENCE OF 25-44.
 RA Paegerstam L.G., Pettersson L.G.;
 RT "The 1,4-Beta-glucan cellobiohydrolases of *Trichoderma reesei* QM
 RT 9414.";
 RL FEBS Lett. 119:97-100 (1980).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 EX MEDLINE=90333255; PubMed=2377893;
 RA Rouvinen J., Berofsky T., Teeri T.T., Knowles J.K.C., Jones T.A.;
 RT "Three-dimensional structure of cellobiohydrolase II from *Trichoderma*
 RT reesei.";
 RL Science 249:380-386 (1990).
 RN [5]
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=97029636; PubMed=8875646;
 RA Koivula A., Reinikainen T., Ruohonen L., Valkesjaervi A.,
 RA Claessens M., Teeman O., Kleywegt G.J., Szardenings M., Rouvinen J.,
 RA Jones T.A., Teeri T.T.;
 RT "The active site of *Trichoderma reesei* cellobiohydrolase II: the role
 RT of tyrosine 169.";
 RL Protein Eng. 9:691-699 (1996).
 CC -!- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
 CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
 CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
 CC (2) EXOCELLULOHYDROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
 CC FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN;
 CC (3) BETA-1,4-GLUCOSIDASES WHICH HYDROLYZE THE CELLULOSE AND OTHER
 CC SHORT CELLO-OLIGOSACCHARIDES TO GLUCOSE.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellulose and cellotetraose, releasing cellobiose from the non-
 CC reducing ends of the chains.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MISCELLANEOUS: T. RESEI PRODUCES TWO DIFFERENT
 CC EXOCELLULOHYDROLASES. THEY ARE UNIQUE IN THAT THEY CAN HYDROLYSE
 CC CRYSTALLINE CELLULOSE IN THE ABSENCE OF ENDOGLUCANASES.
 CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL
 CC HYDROLASES).
 CC -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M16190; AAA34210.1; -;
 CC EMBL; M55080; AAA72922.1; -;
 CC PIR; A26472; A26472.
 CC PIR; A26160; A26160.
 CC PDB; 3CBH; 15-JAN-91.
 CC PDB; 1CB2; 19-MAR-99.
 CC InterPro; IPR000254; CBD fungal.
 CC InterPro; IPR001524; GH 6.
 CC Pfam; PF00734; CBM 1; 1_
 CC Pfam; PF01341; Glyco_hydro 6; 1.
 CC PRINTS; PR00733; GLHYDRLASE6.
 CC ProDom; PD001821; CBD fungal; 1.
 CC ProDom; PD003733; GH_6; 1.
 CC SMART; SM00236; fCBD; 1.
 CC PROSITE; PS00562; CBD FUNGAL; 1.
 CC PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.
 CC PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.
 CC Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal;
 KW 3D-structure. 1 24
 FT SIGNAL 25 471 EXOGLUCANASE II.
 FT CHAIN 25 65 CELLULOSE-BINDING (BY SIMILARITY).
 FT DOMAIN

FT DOMAIN 66 106
 FT DOMAIN 107 471
 FT ACT_SITE 199 245
 FT ACT_SITE 245 245
 FT ACT_SITE 425 425
 FT DISULFID 34 51
 FT DISULFID 45 61
 FT DISULFID 200 259
 FT DISULFID 392 439
 FT CARBOHYD 38 38
 FT CARBOHYD 111 111
 FT CARBOHYD 121 121
 FT CARBOHYD 130 130
 FT CARBOHYD 133 133
 FT CARBOHYD 134 134
 FT CARBOHYD 139 139
 FT CARBOHYD 313 313
 FT CARBOHYD 334 334
 FT MUTAGEN 199 199
 FT MUTAGEN 245 245
 FT CONFLICT 359 359
 FT CONFLICT 449 449
 SQ SEQUENCE 471 AA; 49653 MW; C4711BC335B1BD88 CRC64;
 Query Match 11.6%; Score 226; DB 1; Length 471;
 Best Local Similarity 25.5%; Pred. No. 6.8e-08;
 Matches 94; Conservative 41; Mismatches 156; Indels 78; Gaps 13;
 QY 5 TIASSAILALAVGTETMAHAACSKAYQCGKWDGPTCCESGTCVYDPNPFYGCVP 64
 DB 8 TLATLATLAASVPLEERQA--CSSVMGQCGQNWSGPTCCASGTCV--YSNDYYSOCLP 63
 QY 65 NENLTSTNKSHTTTESAKTTTYSKTTTTEASKTTTTEASKTTTTEASKTTT 124
 DB 64 G--AAASSSTSTRAASTTSVSTSRSS--SATPPPGSTTTTRVPPVSGSTATYSGNPFV 118
 QY 125 TTTKASTSTSSSSSSAS---TNSAVSGGASNGETTRYWDCCKPSCSWPGKADVTSP 180
 DB 119 GVTFWNAVYASEVSSLAISLTCAMATAAAVAK-----VPSFMWLDLTKPL 168
 QY 181 VGSCKDQKTLADNNTONGCVGSS-----YTCDNQPVVVDLADLAYGFAAASISGSEA 235
 DB 169 M-----EQLADIRATANKNGVAGQVYVYDLPRDCAALASNGEY-----SIADGGVA 217
 QY 236 TW-----CCACFELTSTAVKGMVVOVNTGSDLSNTGAHFDLQMPGGVG 286
 DB 218 KYKNYIDTIRQIVVEYSIDIRLLVIEPDSLNLNLNLTGPKCANAAQASAY-----LEC 269
 QY 287 YNGCATQWGAPTDGWRGARYGVGVSSASDCSNLPSALQACKWRFGFKNADNPTM----TY 342
 DB 270 INYAVTQNLNLP-----NVMYLDAGHAGWLGWNPANQDPAAQLFANVY 311
 QY 343 QVTCPKAI 351
 DB 312 KNASSPRAL 320
 RESULT 6
 SIM1 YEAST
 ID SIM1 YEAST STANDARD; PRT; 475 AA.
 AC P40472;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE SIM1 protein precursor.
 GN SIM1 OR PB3 OR YII123W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;

RA Barrell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
RA Churcher C.M., Connor R., Copeley T., Dear S., Devlin K., Fraser A.,
RA Gentles S., Hamlyn N., Horsnell T.S., Hunt S., Jagels K., Jones M.,
RA Louis E., Lye G., Moule S., Moulé T., Odell C., Pearson D.,
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
RA Walsh S.V., Whitehead S.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN (2)
RP IDENTIFICATION.
RX MEDLINE=96155611; PubMed=8574583;
RA Damann C., Diffley J.F.X., Nasmyth K.A.;
RT "S-phase-promoting cyclin-dependent kinases prevent re-replication by
RT inhibiting the transition of replication origins to a pre-replicative
RT state";
RL Curr. Biol. 5:1257-1269(1995).
CC -1- FUNCTION: REQUIRED FOR THE MAINTENANCE OF THE CLB5 KINASE
CC ACTIVITY.
CC -1- SIMILARITY: BELONGS TO THE SUN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z46933; CAA86869.1; --
DR SGD; S0001385; SIM1.
KW Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 475 SIM1 PROTEIN.
FT DOMAIN 58 112 ALA-RICH.
FT DOMAIN 80 203 SER-RICH.
FT CARBOHYD 422 422 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 475 AA; 48070 MW; B7B0F0BC9478612F CRC64;

Query Match 10.1%; Score 198.5; DB 1; Length 475;
Best Local Similarity 23.8%; Pred No. 4e-06;
Matches 115; Conservative 68; Mismatches 152; Indels 149; Gaps 24;

Qy 1 MKFLT-----IASSAILA-----LAVGTEMAHAECISKAY-----YQCGKNWDGPTCCESGS 48
Db 1 MKFSTAVTLISSGAIVSALPHVDVHQDAHQKRAVAYKYVETVVVDSGHTVTPPA 60

Qy 49 -----TCVDYP-----DNPYSOCVFNENLTS 70
Db 61 EVATAATSAIITTSVLAPTSSAAAGIAASIAVSSAALAKNEKISDAASATASTSQGASS 120

Qy 71 TNKSSHKTTTETESAKKTTTGGSKTTTTEASKKTTTTEASK--KTTTTEASKKTTTTK 128
Db 121 SSSSSATSTLESSSVSSSEBAPTSTVVTSSATQSSASATKSSSTSTSPSTSTST 180

Qy 129 KASTSTSSSSASTNYSAVSGGASG-----NGETTRYDCCKPCSCSWPGKADVTSP--V 181
Db 181 TSSTSTSSSSSSSSSSSSSSGGSIYGLDLADFGSPSEKEFGDGTIPCDKFPSSGQVISIDWI 240

Qy 182 GSCNKDKTKLADNTQNCVCGS--SYTCNDNP-----WVSDLLAYGAAASI----- 229
Db 241 GEGGWSGVENTDTSTGGCKEGSYCSYSC---QPGMSKTQW--PSDQPSDGRSVGSLCKN 296

Qy 230 ---SGGSEATWCC-----ACFELFTTFAVK-----GKMWVQV--TNTGSDL----- 267
Db 297 GYLRSNTADADYLCBEGVEAYVVSKLKGVACIKTDPGTENMVIPYVEGGSLPLTV 356

Qy 268 -----GSNTGAHFDLQMPGGVGIYNGCATQWG-----APTQMGARYGGV 308
Db 357 VDQDTYFTWEKTKTAQY--YNNAGVSVEDGCI--WGTSGSIGNWAFLNFGAGSTGV 412

Qy 309 S-----SASDCNLSALQAGCKWRFGWK--NADNPTMTYKQVT 346
Db 413 TVLSLIPNPNNSDALNPNVKIVAADSSNVIGE-----CVYENGEPFGADGCTVS---VT 465

Qy 347 CPKA 350
Db 466 SGKA 469

RESULT 7
MUC1_XENLA
ID MUC1_XENLA STANDARD; PRT; 662 AA.
AC Q05049;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Integumentary mucin C.1 (FIM-C.1) (Fragment).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4; 5; 6 AND 7).
RX TISSUE=Skin;
MEDLINE=93077556; PubMed=1447205;
RA Hauser P., Hoffmann W.;
RT "P-domains as shuffled cysteine-rich modules in integumentary mucin
RT C.1 (FIM-C.1) from Xenopus laevis. Polydispersity and genetic
RT polymorphism";
RL J. Biol. Chem. 267:24620-24624(1992).
CC -1- FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL
CC INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- ALTERNATIVE PRODUCTS: At least 7 isoforms; 1 (shown here), 2, 3,
CC 4, 5, 6 and 7; may be produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: SKIN.
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED.
CC -1- SIMILARITY: CONTAINS 6 P-TYPE (TREFOL) DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; L02115; AAA74725.1; --
DR PIR; A45155; A45155.
DR HSP; P01359; 2PSP.
DR InterPro; IPR000519; P_trefol.
DR Pfam; PF00088; trefol; 6.
DR SMART; SM00018; P; 6.
DR PROSITE; PS00025; P_TREFOL; 6.
KW Repeat; Amphibian skin; Glycoprotein; Alternative splicing.
FT NON TER 1 144
FT DOMAIN 81 144 8 X 8 AA APPROXIMATE TANDEM REPEATS,
FT REPEAT 81 88 ALA/THR-RICH.
FT REPEAT 89 96 1-1.
FT REPEAT 97 104 1-2.
FT REPEAT 105 112 1-3.
FT REPEAT 113 120 1-4.
FT REPEAT 121 128 1-5.
FT REPEAT 129 136 1-6.
FT REPEAT 137 144 1-7.
FT DOMAIN 161 202 P-TYPE 1.
FT DOMAIN 218 301 8 X APPROXIMATE TANDEM REPEATS, THR-RICH.
FT REPEAT 218 224 2-1.
FT REPEAT 225 239 2-2.
FT REPEAT 240 249 2-3.
FT REPEAT 250 259 2-4.
FT REPEAT 260 275 2-5.
FT REPEAT 276 287 2-6.
FT REPEAT 288 294 2-7.
FT REPEAT 295 301 2-8.

FT	DOMAIN	306	347	P-TYPE 2.
FT	DOMAIN	353	353	P-TYPE 3.
FT	DOMAIN	402	522	12 X APPROXIMATE TANDEM REPEATS,
FT	REPEAT	402	411	THR-RICH.
FT	REPEAT	412	419	3-1.
FT	REPEAT	420	431	3-2.
FT	REPEAT	432	443	3-3.
FT	REPEAT	444	453	3-4.
FT	REPEAT	454	460	3-5.
FT	REPEAT	461	472	3-6.
FT	REPEAT	473	479	3-7.
FT	REPEAT	480	491	3-8.
FT	REPEAT	492	498	3-9.
FT	REPEAT	499	515	3-10.
FT	REPEAT	516	522	3-11.
FT	REPEAT	525	566	3-12.
FT	DOMAIN	525	613	P-TYPE 4.
FT	DOMAIN	525	613	P-TYPE 5.
FT	DOMAIN	620	661	P-TYPE 6.
FT	DISULFID	162	188	BY SIMILARITY.
FT	DISULFID	172	187	BY SIMILARITY.
FT	DISULFID	182	199	BY SIMILARITY.
FT	DISULFID	307	333	BY SIMILARITY.
FT	DISULFID	317	332	BY SIMILARITY.
FT	DISULFID	327	344	BY SIMILARITY.
FT	DISULFID	354	380	BY SIMILARITY.
FT	DISULFID	364	379	BY SIMILARITY.
FT	DISULFID	374	391	BY SIMILARITY.
FT	DISULFID	526	552	BY SIMILARITY.
FT	DISULFID	536	551	BY SIMILARITY.
FT	DISULFID	546	563	BY SIMILARITY.
FT	DISULFID	573	599	BY SIMILARITY.
FT	DISULFID	583	598	BY SIMILARITY.
FT	DISULFID	593	610	BY SIMILARITY.
FT	DISULFID	621	647	BY SIMILARITY.
FT	DISULFID	631	646	BY SIMILARITY.
FT	DISULFID	641	658	BY SIMILARITY.
FT	VARSPLIC	240	259	MISSING (IN ISOFORMS 5 AND 6).
FT	VARSPLIC	250	259	MISSING (IN ISOFORMS 4 AND 7).
FT	VARSPLIC	276	294	MISSING (IN ISOFORMS 4 AND 6).
FT	VARSPLIC	278	278	MISSING (IN ISOFORM 5).
FT	VARSPLIC	306	350	MISSING (IN ISOFORMS 2 AND 5).
FT	VARSPLIC	420	498	MISSING (IN ISOFORM 3).
FT	VARIANT	276	276	K -> E.
FT	VARIANT	354	354	C -> R.
FT	VARIANT	415	415	T -> A.
SQ	SEQUENCE	662 AA;	67774 MW;	F085277F1ED2FD40 CRC64;
Query Match				
Best Local Similarity				
Matches 89; Conservative 33; Mismatches 127; Indels 92; Gaps 15;				
QY	32	QCQGNWDGPTCESGST	-----CVDPYDPNPFYSCQVFNENLTSTNKSHTTTTESAKK	86
Db	373	QCQKQK	-----CCFDSISGKWC	-----FYSISQVAATKTTTPTTTTPTTTT 420
QY	87	TTTTKSKKTTTTEASKKTTTTEASKKTTTTEASKKTTTTEASKKTTTTEASKK	TTTTEASKKTTTTEASKKTTTTEASKKTTTTEASKKTTTTEASKKTTTTEASKK	145
Db	421	ATTT--TPITTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTT	TTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTT	478
QY	146	SAVSGGASNGETTRYDCKPSCSWP	-----GKADVTSPVSGCNKDGKTLAD	193
Db	479	TKATTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTT	TTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTTPTTTT	538
QY	194	-----NNTQNGCVGGSSVT	-----CNDNQFWVDDLAYGFAAASISGGSEATWC	239
Db	539	PGITESQCRSGKCCFDSIPQTKWCFYSLPQVADCKVAPSSRVDCFGGITADQCRQNC	598	
QY	240	ACELFTSTAVKGGKVVQVNTGSDLSNGAHDLOMPGGGVGVIYNGCATQWCAPTD	299	
Db	599	CFD---7-SSISGTKWCFYSTSQGNMCS	-----GPPTK	627
QY	300	GWGARYGGVSSASDCNLSALQAGCKWRFQWPKNADNPTM	340	

Db	628	RRDCGPGISS-SVCIN	-----RGCCW	-----DNSVM	653
RESULT 8					
GUNB_FUSOX STANDARD; PRT; 462 AA.					
ID	GUNB_FUSOX	STANDARD;	PRT;	462	AA.
AC	P46236;				
DT	01-NOV-1995	(Rel. 32, Created)			
DT	01-NOV-1995	(Rel. 32, Last sequence update)			
DT	01-NOV-1995	(Rel. 32, Last annotation update)			
DE	Putative endoglucanase type B precursor (EC 3.2.1.4)	(Endo-1,4-beta-glucanase) (Cellulase).			
OS	Fusarium oxysporum.				
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;				
OC	Hypocreales; Mitosporic Hypocreales; Fusarium.				
OX	NCBI_TaxID=5507;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=95047531; PubMed=7959045;				
RA	Sheppard P.O., Grant F.J., Cort P.J., Sprecher C.A., Foster D.C.,				
RA	Hagen F.S., Upshall A., McKnight G.L., O'Hara P.J.;				
RT	"The use of conserved cellulase family-specific sequences to clone				
RT	cellulase homologue cDNAs from Fusarium oxysporum.";				
RL	Gene 150.163-167(1994).				
CC	-1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic				
CC	linkages in cellulose.				
CC	-1- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).				
CC	-1- SIMILARITY: BELONGS TO CELLULASE FAMILY B (FAMILY 6 OF GLYCOSYL				
CC	HYDROLASES).				

This SWISS-PROT entry is copyright. It is produced through a collaboration					
between the Swiss Institute of Bioinformatics and the EMBL outstation -					
the European Bioinformatics Institute. There are no restrictions on its					
use by non-profit institutions as long as its content is in no way					
modified and this statement is not removed. Usage by and for commercial					
entities requires a license agreement (See http://www.isb-sib.ch/announce/					
or send an email to license@isb-sib.ch).					

EMBL; L29377; AAA65585.1; -					
HSSP; P07987; 1CB2.					
InterPro; IPR000254; CBD fungal.					
InterPro; IPR001524; GH 6.					
Pfam; PF00734; CBM_1; 1.					
Pfam; PF01341; Glyco_hydro_6; 1.					
PRINTS; PR00733; GLHYDRLASE6.					
ProDom; PD001821; CBD fungal; 1.					
ProDom; PD003733; GH 6; 1.					
SMART; SM00236; fCBD; 1.					
PROSITE; PS00562; CBD_FUNGAL; 1.					
PROSITE; PS00655; GLYCOSYL_HYDROL_F6_1; 1.					
PROSITE; PS00656; GLYCOSYL_HYDROL_F6_2; 1.					
KW	Cellulose degradation; Hydrolase; Glycosidase; Glycoprotein; Signal.				
FT	SIGNAL	1	16	POTENTIAL.	
FT	CHAIN	17	462	PUTATIVE ENDOGLUCANASE TYPE B.	
FT	DOMAIN	17	65	CELLULOSE-BINDING.	
FT	DOMAIN	66	99	LINKER.	
FT	DOMAIN	100	462	CATALYTIC.	
FT	ACT SITE	190	190	BY SIMILARITY.	
FT	ACT SITE	236	236	PROTON DONOR (BY SIMILARITY).	
FT	ACT SITE	416	416	NUCLEOPHILE (BY SIMILARITY).	
FT	DISULFID	33	50	BY SIMILARITY.	
FT	DISULFID	44	60	BY SIMILARITY.	
FT	DISULFID	191	250	BY SIMILARITY.	
FT	DISULFID	383	430	BY SIMILARITY.	
FT	CARBOHYD	37	37	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	223	223	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	272	272	N-LINKED (GLCNAC. . .) (POTENTIAL).	
FT	CARBOHYD	317	317	N-LINKED (GLCNAC. . .) (POTENTIAL).	
SQ	SEQUENCE	462 AA;	49207 MW;	E25B2F5B828B637F CRC64;	
Query Match					8.8%; Score 172.5; DB 1; Length 462;
Best Local Similarity					30.9%; Pred. No. 0.00019;

DR Pfam; PF00704; Glyco_hydro_18; 1;
DR PROSITE; PS01095; CHITINASE_18; 1;
KW Hydrolase; Glycosidase; Chitin degradation; Signal; Glycoprotein;
KW Chitin-binding; Multigene family.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 567 CHITINASE 3.
FT ACT_SITE 157 157 PROTON DONOR (BY SIMILARITY).
FT CARBOHYD 159 159 N-LINKED (GLCNAC...) (POTENTIAL).
FT DOMAIN 319 436 SER/THR-RICH.
SQ SEQUENCE 567 AA; 60060 MW; DD843126F6E22C2 CRC64;

Query Match 8.6%; Score 169; DB 1; Length 567;
Best Local Similarity 27.5%; Pred. No. 0.00038;
Matches 57; Conservative 34; Mismatches 74; Indels 42; Gaps 6;

QY 66 ENLTS-TNKSSHKTTTTSASKKTTTTSKSKTTTTEASKKTTTTEASKKTTTTEASKKTTT 124
DB 304 ENKAIKVKKSGEETSSSTTTTTT-----TTSTISSSSSSSKTSKTSKTSSTSSSIS 358
QY 125 TTTKASTSTSSSSASSTNYSVGGASGNGETTRYWDCKKPCSCWFGKADVTSPVGS 184
DB 359 STTSTSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSST 401
QY 185 NKDGKTLADNNNTQNGCVGSSVTCNDNQPVVDDLAYGFAAASISGGSEATWCCACFEL 244
DB 402 SLSSSTI---STASSTSTTSVTSSTETVPVTPSSLS---SAITIPGDS----- 444
QY 245 TPTSTAVKGMVQVNTGSDLGST 271
DB 445 ---TTTGISKSSSTKPKATSTTSALSST 469

RESULT 11
VGLX_HSVB STANDARD; PRT; 797 AA.
AC P28968;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Glycoprotein X precursor.
GN 71.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1";
RL Virology 189:304-316(1992).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M86664; AAB02506.1; --
DR PIR; H36802; VGBEX1.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 797 GLYCOPROTEIN X.
FT DOMAIN 23 465 SER/THR-RICH.
FT TRANSMEM 766 790 POTENTIAL.
FT CARBOHYD 590 590 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 797 AA; 80342 MW; 50C9ED9211F5E5B2 CRC64;

Query Match 8.4%; Score 164; DB 1; Length 797;
Best Local Similarity 24.2%; Pred. No. 0.0011;
Matches 64; Conservative 36; Mismatches 127; Indels 38; Gaps 6;

QY 70 STNKSHKTT---TTBSAKKTTTTSKSKTTTTEASKKTTTTEASKKTTT-----TEAS 120
DB 46 TTNSSSSPTTSPPTTSSPSTSTHTSSPSTSTQSSSTAATSSAPSTASTTSIPST 105
QY 121 KKTITTTKASTSTSSSSSAST---NYSVSGASGNGETTRYWDC--KPCSCWPGKA 175
DB 106 TEITTTTPTASTTTTPTTAAPTAAATTAATTAATTAATTAATTAATTAATTAATTA 165
QY 176 DVTSPVGSCKNGKTLADNNNTQNGCVGSSVTCNDNQPVVDDLAYGFAAASISGGSEA 235
DB 166 TTTTATTTTPTASTTTTDTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 213
QY 236 TWCCACFELTFTAVKGMVQVNTGSDLGSTCAHFDLQMPGGVGIYNGCATQWG 295
DB 214 TTTAAATTAATTSATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 261
QY 296 APTDGMGARYGVGSSASDCSNLPSA 320
DB 262 SPTSGSTSTTGASTSTPSASTATSA 286

RESULT 12
GUN2_TRIE STANDARD; PRT; 418 AA.
AC P07982;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Endoglucanase EG-II precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase)
DE (Cellulase).
GN EGL2 OR EGLII.
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VTT-D-80133;
RX MEDLINE=8825850; PubMed=3384334;
RA Saloheimo M., Lehtovaara P., Penttilae M., Teeri T.T., Staahlberg J.,
RA Johansson G., Pettersson G., Clayssens M., Tomme P., Knowles J.K.C.;
RT "EGII, a new endoglucanase from Trichoderma reesei: the
RT characterization of both gene and enzyme";
RL Gene 83:11-21(1988).
RN [2]
RP ACTIVE SITE GLU-350.
RX MEDLINE=93131031; PubMed=8093602;
RX Macarion R., van Beeumen J., Henrissat B., de la Mata I.,
RA Clayssens M.;
RT "Identification of an essential glutamate residue in the active site
RT of endoglucanase II from Trichoderma reesei";
RL FEBS Lett. 316:137-140(1993).
CC -!- FUNCTION: THE BIOLOGICAL CONVERSION OF CELLULOSE TO GLUCOSE
CC GENERALLY REQUIRES THREE TYPES OF HYDROLYTIC ENZYMES:
CC (1) ENDOGLUCANASES WHICH CUT INTERNAL BETA-1,4-GLUCOSIDIC BONDS;
CC (2) EXOCELLULOHYDROLASES THAT CUT THE DISSACCHARIDE CELLULOSE
CC (3) FROM THE NONREDUCING END OF THE CELLULOSE POLYMER CHAIN.
CC SHORT CELLO-OLIGOSACCHARIDES WHICH HYDROLYZE THE CELLULOSE AND OTHER
CC CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -!- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -!- SIMILARITY: CONTAINS 1 FUNGAL-TYPE CELLULOSE-BINDING DOMAIN (CBD).
CC -!- CAUTION: WAS ORIGINALLY CALLED ENDOGLUCANASE EG-III.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC

Db	1205	-----TNTGAETTTTGAETKVTVTSSISRFNHAETQASATDVIG-HSSSV 1252
Qy	309	SSASDCSNLPSALQAG 324
Db	1253	VSVSETGNTKSLTSSG 1268
RESULT 14		
Y899 CABEL		
ID	YS89 CABEL	STANDARD; PRT: 3178 AA.
AC	Q09624; Q09625; Q09624;	
DT	01-NOV-1995 (Rel. 32, Created)	
DT	15-JUN-2002 (Rel. 41, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Hypothetical protein ZK945.9 in chromosome II.	
GN	ZK945.9/ZK945.10.	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;	
OC	Rhabditidae; Peloderinae; Caenorhabditis.	
OX	NCBI_TaxID=6239;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Bristol N2;	
RA	Wilkinson-Sproat J.;	
RL	Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	REVISIONS.	
RA	Durbin R.;	
RL	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.	
CC	-1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).	
CC	-1- SIMILARITY: CONTAINS 1 GPS DOMAIN.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	[1]	
CC	EMBL; Z48544; CAB70192.1; JOINED.	
DR	EMBL; Z48582; CAB70192.1; JOINED.	
DR	EMBL; Z48582; CAB70201.1; JOINED.	
DR	EMBL; Z48544; CAB70201.1; JOINED.	
DR	WormPep; ZK945.9; CB25697.	
DR	InterPro; IPR002111; Cat channel TrpL.	
DR	InterPro; IPR001024; Lipoxigenase_LH2.	
DR	InterPro; IPR000636; M-channel_nlg.	
DR	InterPro; IPR000203; PKD_cys_rich.	
DR	Pfam; PF00520; Igm trans; 1.	
DR	Pfam; PF01477; PLAT; 1.	
DR	Pfam; PF01825; GPS; 1.	
DR	SMART; SM00303; GPS; 1.	
DR	SMART; SM00308; LH2; 1.	
KW	Hypothetical protein; Transmembrane.	
FT	DOMAIN 266 1196	SER/THR-RICH.
FT	DOMAIN 1105 1241	GLY/SER-RICH.
FT	DOMAIN 2071 2120	GPS.
FT	TRANSMEM 13 30	POTENTIAL.
FT	TRANSMEM 51 73	POTENTIAL.
FT	TRANSMEM 2139 2161	POTENTIAL.
FT	TRANSMEM 2348 2367	POTENTIAL.
FT	TRANSMEM 2390 2412	POTENTIAL.
FT	TRANSMEM 2451 2468	POTENTIAL.
FT	TRANSMEM 2483 2505	POTENTIAL.
FT	TRANSMEM 2567 2589	POTENTIAL.
FT	TRANSMEM 2836 2858	POTENTIAL.
FT	TRANSMEM 2939 2961	POTENTIAL.
FT	TRANSMEM 2976 2998	POTENTIAL.
FT	TRANSMEM 3038 3060	POTENTIAL.
SQ	SEQUENCE 3178 AA; 344726 MW; F8239436D03666CD CRC64;	
	8.3%; Score 162; DB 1; Length 3178;	
Query Match		

Best Local Similarity	28.7%;	Pred. No. 0.0056;
Matches	79;	Conservative 38; Mismatches 130; Indels 28; Gaps 12;
Qy	42	TCCSGSTCVDPYDPNPFYSCVQPNENLTSTNKS- --HKTTTBSAKTTTGGSKTTT 98
Db	365	TPTEITSTLSSLPDNAICSYL- -DETTTSTFTTMTLTSTTTEPSTSTTTTTEVTSST 422
Qy	99	TEASKKTTT-TEASKKTTTTEASKKTTTTEASKK- -TSTSSSSSSASTNYSAVSG- -GAS 153
Db	423	VTTTEPTTTLTSTASTSTTEPSTSTVTTPSTSTVTTPSTSTVTSSSSSTVTTPSTSTST 482
Qy	154	GNGETRYWDCCKPSCSWPGKADVTSPVGSCKDKGLADNNTQNGCVGSSSYTCNDNQP 213
Db	483	TSPSTVTVTSTTAPSTSTTGFSSSSSTPSSSTASSSVSTASTOSTSTTQSSSTTKSET 542
Qy	214	WVVSDDLAYGFAAASISGSEATWCCACFELTFTS- -TAVKGGKMWVQVTN- -TGSDLGNT 271
Db	543	TTSDDGTNPDF- -YFVEKATTTFYDSTSVNLTLNGLGIIGVQTSIECTSPSTSSNVVSTT 600
Qy	272	--GAHF- -DLQMP- -GG-----GVGIYNGCAT 292
Db	601	KDCACTKSVSMPLRGTYPASTFVPGPGNYTFRAT 635
RESULT 15		
STUB DROME		
ID	STUB DROME	STANDARD; PRT: 786 AA.
AC	Q05319;	
DT	01-JUN-1994 (Rel. 29, Created)	
DT	01-JUN-1994 (Rel. 29, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Serine proteinase stubble (EC 3.4.21.-) (Stubble-stubblويد protein).	
GN	SB OR SB-SB.	
OS	Drosophila melanogaster (fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;	
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
OC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Oregon-R;	
RA	Medline=93281671; PubMed=7685111;	
RA	Appel L.F., Prout M., Abu-Shumays R., Hammonds A., Garbe J.C.,	
RT	Fristrom D., Fristrom J.;	
RT	"The drosophila stubble-stubblويد gene encodes an apparent	
RT	transmembrane serine protease required for epithelial	
RT	morphogenesis.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 90:4937-4941(1993).	
CC	-1- FUNCTION: HORMONE DEPENDENT PROTEASE REQUIRED FOR EPITHELIAL	
CC	MORPHOGENESIS. HAS A DUAL FUNCTION, DETACHES IMAGINAL DISC CELLS	
CC	FORM EXTRACELLULAR MATRICES THROUGH ITS EXTRACELLULAR PROTEOLYTIC	
CC	DOMAIN AND TRANSMITS AN OUTSIDE-TO-INSIDE SIGNAL TO ITS	
CC	INTRACELLULAR DOMAIN TO MODIFY THE CYTOSKELETON DURING	
CC	MORPHOGENESIS..MAY BE ABLE TO ACTIVATE ITSELF.	
CC	-1- SUBCELLULAR LOCATION: TYPE II membrane protein.	
CC	-1- INDUCTION: BY 20-HYDROXYECYDSONE (20HE).	
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1.	
CC	-1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-24 IS THE INITIATOR.	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	[1]	
CC	EMBL; L11451; AAA28918.1; -	
DR	HSSP; P00763; IDPO.	
DR	MEROPS; S01.225; -	
DR	FlyBase; FBgn0003319; Sb.	
DR	InterPro; IPR001314; Chymotrypsin.	
DR	InterPro; IPR001254; Ser_protease_Try.	

Search completed: June 18, 2003, 15:32:05
Job time : 11.4426 secs

THIS PAGE BLANK (USPTO)